MSUHEP-100421, SMU-HEP-10-07, arXiv:1004.4624 [hep-ph]

Uncertainty induced by QCD coupling in the CTEQ global analysis of parton distributions

Hung-Liang Lai,[1,2] Joey Huston,[2] Zhao Li,[2] Pavel Nadolsky,[3]
Jon Pumplin,[2], Daniel Stump,[2] and C.-P. Yuan[2]

[1] *Taipei Municipal University of Education, Taipei, Taiwan*
[2] *Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824-1116, U.S.A.*
[3] *Department of Physics, Southern Methodist University, Dallas, TX 75275-0175, U.S.A.*

(Dated: August 19, 2010)

Abstract

We examine the dependence of parton distribution functions (PDFs) on the value of the QCD coupling strength $\alpha_s(M_Z)$. We explain a simple method that is rigorously valid in the quadratic approximation normally applied in PDF fitting, and fully reproduces the correlated dependence of theoretical cross sections on α_s and PDF parameters. This method is based on a statistical relation that allows one to add the uncertainty produced by α_s, computed with some special PDF sets, in quadrature with the PDF uncertainty obtained for the fixed α_s value (such as the CTEQ6.6 PDF set). A series of four CTEQ6.6AS PDFs realizing this approach, for α_s values in the interval $0.116 \leq \alpha_s(M_Z) \leq 0.120$, is presented. Using these PDFs, the combined α_s and PDF uncertainty is assessed for theoretical predictions at the Fermilab Tevatron and Large Hadron Collider.

Contents

I. INTRODUCTION

The global analysis of quantum chromodynamics (QCD) refers to the use of data from many short-distance scattering processes to construct, within some approximations, universal parton distribution functions (PDFs) that can be used to calculate hadronic cross sections for the QCD and electroweak theories. An important part of the analysis is to determine the *uncertainties* of the PDFs. The uncertainties have several sources, both experimental and theoretical. An example of an uncorrelated experimental uncertainty, i.e., an uncertainty induced by *data*, is the statistical error on the cross section measurements used as the input to the global analysis. An example of a correlated experimental uncertainty, also induced by *data*, is the luminosity error, which modifies the normalization of all cross sections measured by one experiment by the same factor. An example of a theoretical uncertainty is the choice of a momentum scale Q used in the theoretical calculations for some process; this uncertainty depends on the order of perturbative approximation (next-to-leading order, NLO, in α_s for the CTEQ analysis in this paper) and is highly correlated between different points in the phase space of the process. Methods for calculating these various uncertainties have been developed for the CTEQ analysis [1] and other analyses of PDFs [2–6].

Two overall theoretical uncertainties of the global analysis are the "parametrization error" and the uncertainty in the QCD coupling strength α_s. The first concerns the parametric form for the nonperturbative PDFs at a low momentum scale Q_0, used as a boundary condition for predicting the PDFs at other scales through their perturbative evolution. The choice of the functional form affects the predictions and so contributes an uncertainty to the published PDFs.

The strong coupling α_s is another basic parameter that affects every theoretical cross section in the global fit. As with the PDF parametrizations, one must decide how to evaluate the scale dependence of the running coupling constant $\alpha_s(Q)$ and provide a value of α_s at some scale Q' as a boundary condition. Both choices contribute additional uncertainties to the determination of PDFs. The Q dependence of $\alpha_s(Q)$ is not uniquely determined, but estimated from the NLO approximation of the renormalization group (RG) equation for $\alpha_s(Q)$:

$$Q\frac{d\alpha_s}{dQ} = -\frac{\beta_0}{2\pi}\alpha_s^2 - \frac{\beta_1}{8\pi^2}\alpha_s^3 + O(\alpha_s^4), \tag{1}$$

where $\beta_0 = 11 - (2/3)n_f$, and $\beta_1 = 102 - (38/3)n_f$. In NLO perturbation theory, we neglect terms of order α_s^4. Then there are different solutions of (1), formally equivalent in the NLO approximation, but differing in higher orders of perturbation theory. Previous studies have shown that the associated uncertainty, coming from the approximation for the Q dependence of $\alpha_s(Q)$, is small compared to other sources of the PDF uncertainty [7].

In this paper we are concerned with the value of α_s at a typical scale (Q) for hard interactions. The differential equation (1) requires an initial value, commonly taken from a combination of precise experimental measurements. The conventional choice is the value of $\alpha_s(Q)$ at the Z boson mass, $Q = M_Z$. Any value of Q could be taken to set the initial value, but $Q = M_Z$ is the natural choice, because M_Z is known precisely, and some experiments determine the strong coupling directly at this momentum scale.

The input value of $\alpha_s(M_Z)$ and its uncertainty are continually being refined, as new experimental data are obtained. The global PDF analysis itself could "determine" (i.e., estimate within the errors) the value of $\alpha_s(M_Z)$ from the comprehensive hadronic data it

examines. However, the resulting determination turns out to be quite uncertain: the included scattering processes (deep-inelastic scattering, Drell-Yan process, $p\overline{p} \to$ jets, etc.) have limited accuracy for measuring $\alpha_s(M_Z)$, precisely because they depend on uncertain PDFs.

The most accurate value of $\alpha_s(M_Z)$ is instead available from a worldwide compilation provided by the Particle Data Group (PDG) [8, 9]. The 2009 world average value of α_s based on eight different measurement techniques is [8]

$$\alpha_s(M_Z) = 0.1184 \ \pm \ 0.0007. \tag{2}$$

Most experimental data that are an input to this world average are independent of parton distributions. For example, precision data from $e^+ e^-$ annihilation at LEP provide strong and direct constraints on $\alpha_s(M_Z)$, independently of the PDFs. The uncertainty range in Eq. (2) nominally corresponds to the 68% confidence level (CL); yet, when assigning this level, one must be aware that the analysis mixes NLO and NNLO calculations in various techniques and relies on some theoretical models. Even with these caveats, the accuracy of the world-average value is clearly superior to the value determined solely from the hadronic scattering data.

Traditionally, our analysis presents the *best-fit* PDFs and their parametrization uncertainties for a constant value of $\alpha_s(M_Z)$ close to its latest world-average central value. Separately, the uncertainty in the PDFs induced by the uncertainty in $\alpha_s(M_Z)$ is assessed, by producing a few *alternative PDF fits* for a range of values of $\alpha_s(M_Z)$ around the central $\alpha_s(M_Z)$ value [7]. We add the two uncertainties in quadrature to estimate the total uncertainty resulting both from the PDF parametrization and α_s. In this paper, we examine this procedure (designated as "fitting method 1") in detail, in order to determine how well it captures the correlation, or *interplay*, of the α_s uncertainty and PDF uncertainty.

On the other hand, the most complete method to evaluate the combined PDF + α_s uncertainty in the global fit is to vary the theoretical value of $\alpha_s(M_Z)$ as an additional fitting parameter, while *including* the world-average value of $\alpha_s(M_Z)$ with its experimental uncertainty as a precise experimental constraint on $\alpha_s(M_Z)$, in combination with the rest of the hadronic data. The uncertainty from α_s can then be deduced by this technique – designated as "fitting method 2" – on the same footing, and by the same Hessian techniques, as the uncertainties in the PDF parameters.[1]

Our main finding is that the simpler framework of the fitting method 1 (based on the α_s series) is sufficient to reproduce *all* correlated dependence of α_s and PDF parameters arising in the fuller treatment of the fitting method 2. Consequently, the α_s series of CTEQ6.6 fits, generated in fitting method 1, is sufficient for determining the total PDF+α_s uncertainty in theory calculations, e.g., predictions for the Large Hadron Collider (LHC). This conclusion is borne out both numerically and formally, within the quadratic approximation for the log-likelihood function χ^2 in the vicinity of the best fit. The total PDF + α_s uncertainty can be thus calculated without resorting to the alternative, quite elaborate, methods [6, 10, 11], but simply by computing the theoretical cross sections for two additional CTEQ6.6 PDF sets for the extreme values of α_s, besides the usual cross sections for the 44 CTEQ6.6 eigenvector PDFs.

Our conclusion that the addition in quadrature is sufficient, regardless of the magnitude of the correlation, follows from an easily overlooked distinction between the "absence of correlation", on one hand, and "the independent addition of the α_s uncertainty and PDF uncertainty", on the other hand. As we show, variations in α_s generally induce compensating

adjustments in the preferred PDF parameters (correlation) to preserve agreement with those experimental data sets that simultaneously constrain α_s and the PDFs. At the same time, it is possible to define an "α_s uncertainty" that quantifies all correlation effects, is independent from the PDF uncertainty for the best-fit α_s, and can be added in quadrature.

Section II presents the new α_s series of PDFs, based on the 2009 world average value of $\alpha_s(M_Z) = 0.118$ and four additional PDF sets found for $0.116 \leq \alpha_s \leq 0.120$. This series, denoted as CTEQ6.6AS, extends the CTEQ6.6 series of PDFs [12]; it uses the same input data and methods as the CTEQ6.6 study. In the same section, we also introduce an alternative fit, designated CTEQ6.6FAS, with a floating α_s, constrained by the world-average value as an experimental input. We provide a heuristic explanation of why the CTEQ6.6AS and CTEQ6.6FAS methods should result in close estimates of the total PDF + α_s uncertainty. The detailed proof is given in the Appendix. In Section III, we compare the CTEQ6.6AS and CTEQ6.6FAS methods numerically. Also, applications based on the resulting PDFs are shown, including our estimates for the current α_s uncertainty on select theoretical predictions for the LHC.

Section IV addresses a separate issue, the degree of the correlation between the α_s and PDF uncertainties imposed by the hadronic data in the current fits. We elucidate this issue by applying the correlation analysis of Refs. [1, 12, 13] to explore which data sets in the CTEQ6.6 sample impose the tightest constraints on α_s. The paper concludes in Section V with our recommendation to use the α_s series, CTEQ6.6AS, combined in quadrature with CTEQ6.6, to assess the effect of α_s uncertainty on theoretical predictions.

II. QCD COUPLING AS AN INPUT AND OUTPUT OF THE PDF ANALYSIS

A. Diagonalization of the error matrix with respect to α_s

Perturbative QCD cross sections calculated to next-to-leading order (NLO) have uncertainties arising from (1) the residual scale dependence of the cross section due to the truncation of the series at NLO, (2) the PDF uncertainties derived from the global PDF fits and (3) the uncertainty in the value of $\alpha_s(M_Z)$ used in the cross section calculation. The residual scale dependence will depend on the individual perturbative cross section being evaluated. The PDF uncertainty can be determined using the PDF eigenvector sets provided by the global fitting groups. CTEQ and MSTW provide eigenvector PDFs (44 for CTEQ6.6 [12] and 40 for MSTW2008 [10]) determined using the Hessian method, while the NNPDF collaboration provides $\sim 1,000$ PDFs [6]. The value of $\alpha_s(M_Z)$ used in any cross section evaluation must be the same as its value in the global fit providing the PDFs. Thus, the uncertainty in $\alpha_s(M_Z)$ must be evaluated by using the PDF sets in which the same value of $\alpha_s(M_Z)$ has been assumed.

The treatment of the α_s uncertainties in any NLO cross section evaluation depends on their relative size compared to the intrinsic PDF uncertainties, and on whether there exists a correlation between the value of $\alpha_s(M_Z)$ and the Hessian error PDFs. One cannot exclude the possibility that the PDF and α_s uncertainties affect each other. For example, it has been known for a long time that the shape of the gluon PDF found from the global fit changes considerably if $\alpha_s(M_Z)$ is changed in the fit.

There are two prevailing approaches to the choice of the value of $\alpha_s(M_Z)$ in the global PDF fits: either the world average is taken as an *input*, or $\alpha_s(M_Z)$ is determined as an *output*

of the global fit. The CTEQ, HERAPDF and NNPDF sets of PDFs use the first approach, while MSTW uses the second. Related to this choice is the evaluation of the uncertainty in $\alpha_s(M_Z)$: either the uncertainty can be directly related to that determined by the world average, or the uncertainty can be determined by examining the impact of varying $\alpha_s(M_Z)$ in the global fit. Again, CTEQ and the NNPDF collaborations have used the first approach, and MSTW the second.

The reason behind the first approach is that the most precise processes contributing to the world average (τ and quarkonium decays and e^+e^- event shapes at the Z-pole energy) are free of the PDF uncertainties and place stronger constraints on the value of $\alpha_s(M_Z)$ than the hadronic scattering data alone. Knowledge gained from those processes, summarized in the most recent PDG value of $\alpha_s(M_Z)$, naturally provides a useful external input to the global QCD analysis. After all, the idea of *global* analysis is that QCD is a fundamental theory that describes all aspects of strong interactions. The value of $\alpha_s(M_Z)$ is universal. The value that we use should be the most accurately determined value.

The second approach has the advantage of being independent from other sources. It circumvents the issue of combining constraints on $\alpha_s(M_Z)$ from heterogeneous measurements and mixing different theoretical frameworks in the combination procedure. The price to pay, however, is the loss of the constraining power supplied by the most precise measurements.

As an example of the uncertainty in $\alpha_s(M_Z)$, consider again the 2009 world average [8] in Eq. (2), its determination is dominated by the processes that do not require the PDFs. The interpretation of the confidence level assigned to this uncertainty (nominally 68%) is still not fully settled for the reasons mentioned in Section I. For the purposes of this study, we adopt a somewhat more conservative estimate of the uncertainty proposed at the 2009 Les Houches workshop [14] and adopted by the PDF4LHC working group [15],

$$\alpha_s(M_Z) = 0.118 \pm 0.002 \qquad (90\%\ \text{CL}), \tag{3}$$

corresponding to an uncertainty of ± 0.0012 at the 68% CL.

Compare that to the value obtained by MSTW from fitting α_s in the PDF analysis [3, 10],

$$\alpha_s(M_Z) = 0.1202 \ \begin{cases} +0.0012 \\ -0.0015 \end{cases} \qquad (68\%\ \text{CL}), \tag{4}$$

$$\alpha_s(M_Z) = 0.1202 \ \begin{cases} +0.0039 \\ -0.0034 \end{cases} \qquad (90\%\ \text{CL}) \tag{5}$$

at next-to-leading order; or

$$\alpha_s(M_Z) = 0.1171 \pm 0.0014 \qquad (68\%\ \text{CL}) \tag{6}$$

at next-to-next-to-leading order. Or, compare that to our own determination of $\alpha_s(M_Z) = 0.118 \pm 0.005$ (90% CL) from a fit without the world-average constraint that is described below, cf. Eq. (13). The result is consistent with that in Eq. (2); but the uncertainty on $\alpha_s(M_Z)$ is considerably larger if α_s is fitted without the world-average constraint. The central value of $\alpha_s(M_Z)$ returned by this fit is 0.118, which coincides with the fixed value used in the CTEQ6.6 PDF analysis.

One can also envision a third, most general approach, in which the $\alpha_s(M_Z)$ range measured by the most precise techniques, i.e. the world average value, is included as an *input* together

with the usual hadronic scattering data; and the theoretical value of $\alpha_s(M_Z)$, fitted as a free parameter together with the PDF parameters, is returned as an *output* constrained by the combination of *all* measurements. Such a fit is the most direct in quantifying improvements in the accuracy of the output $\alpha_s(M_Z)$ resulting from the hadronic scattering data (as compared to the more precise constraints); as well as in probing the correlation between $\alpha_s(M_Z)$ and PDF parameters.

In this paper, we examine the CTEQ global hadronic data and the world-average $\alpha_s(M_Z)$ in Eq. (3) according to this more general approach, and apply the usual Hessian technique [1] to study the combined PDF + α_s uncertainty (i.e. the usual PDF shape parameters plus 1 more parameter for α_s). However, when implemented straightforwardly, this approach runs into a practical inconvenience: each of the eigenvector PDFs is associated with its own value of $\alpha_s(M_Z)$.

The quadratic approximation provides a remarkable bypass for this shortcoming. By applying the Data Set Diagonalization method introduced in Ref. [17], it is possible to re-diagonalize the parameter space so that only one eigenvector corresponds to the change of α_s, whereas all the other eigenvectors are immune to the change in α_s. (As shown in Ref. [17], this conclusion also holds for any other distinct parameter in the global analysis.) In other words, it is easy to construct a pair of additional PDF sets corresponding to the maximal tolerated excursions of α_s. The α_s uncertainty computed from the difference of these PDFs can be added in quadrature to the CTEQ6.6 uncertainty to reproduce the combined uncertainty with the full PDF-α_s correlation. We refer to the reduced computation as "fitting method 1," and the full computation as "fitting method 2."

A formal proof that methods 1 and 2 are equivalent, within the quadratic approximation, is given in the Appendix. Here we provide a heuristic argument that illustrates the choice of the two PDF eigenvector sets probing the α_s uncertainty.

The α_s parameter in the full fit is associated with a new twenty-third direction that is orthogonal to the hyperplane of the PDF parameters for the central value of $\alpha_s(M_Z) = 0.118$; *i.e.*, the 22-parameter space probed by the CTEQ6.6 PDF set. Originally, the orthonormal eigenvector PDFs of the CTEQ6.6 fit correspond to a representation such that the excursion by the same distance from the best fit in the 22-dimensional hyperspace results in the same increase of χ^2; that is, the surfaces of constant χ^2 are 22-dimensional hyperspheres.

If α_s is allowed to deviate from its best-fit value, a preferred direction emerges in the 22-dimensional hyperspace, along which $\Delta\chi^2$ grows most slowly, because changes in α_s are compensated by changes in some linear combination of the PDF parameters. Rotate the CTEQ6.6 basis so that the "physical" PDF parameter a_1 corresponds to the variation of this preferred combination from its best-fit value, and analogously, the "physical" parameter a_0 corresponds to the variation of $\alpha_s(M_Z)$ from its best-fit value (which remains at 0.118). With this choice, $a_0 = 0$ and $a_1 = 0$ at the best fit. All correlations between α_s and PDFs are encapsulated in the χ^2 dependence on a_0 and a_1. Let us focus on this dependence. Uncertainties due to the other (combinations of) PDF parameters a_2, ..., a_{22} are independent of α_s and can be added in quadrature at the end.

In the $\{a_0, a_1\}$ plane, the set of the allowed PDFs corresponds to the inside of an ellipse $\Delta\chi^2 \leq T^2$, where T is the tolerance parameter or some other parameter defining the typical χ^2 at the boundary of the allowed region. The CTEQ6.6 fit is equivalent to probing the inside of the ellipse for a fixed $a_0 = 0$, as shown in the left inset of Fig. 1. The contribution

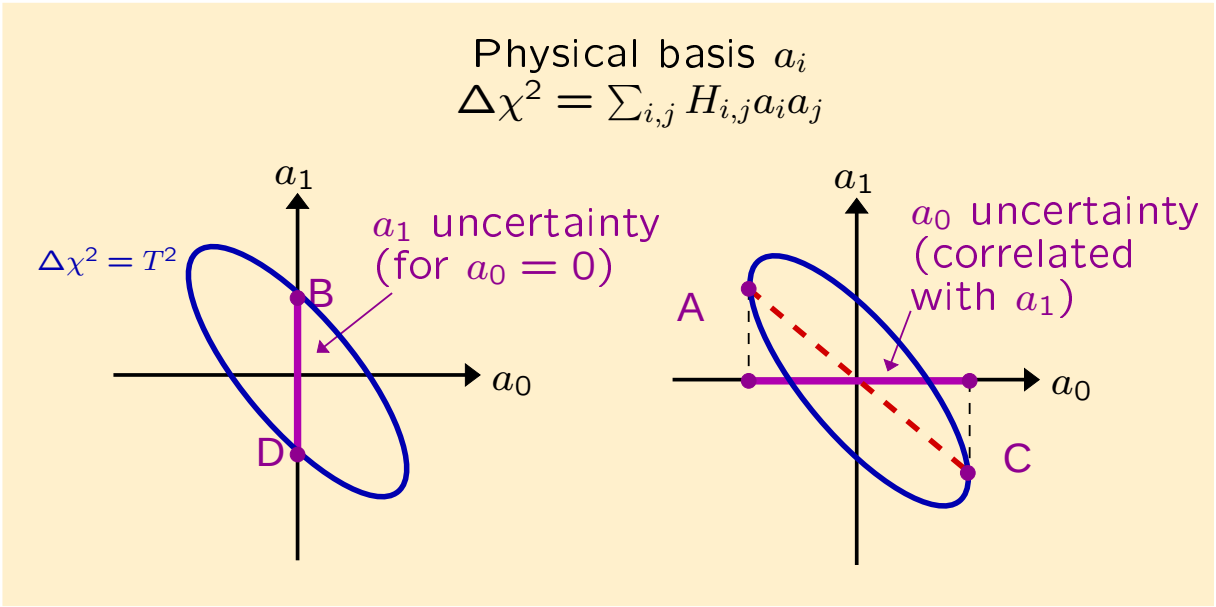

FIG. 1: Imagine fitting two free parameters. The left figure shows uncertainties computed in a fit with $a_0 = 0$ and free a_1. The right figure shows a χ^2 scan over a_0 with free a_1.

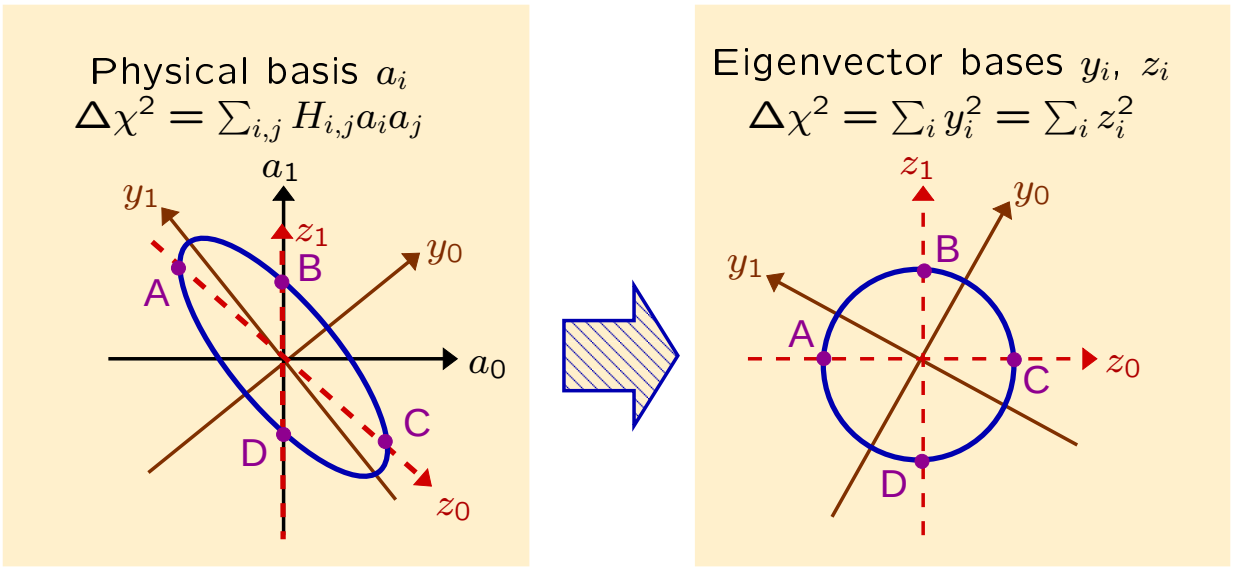

FIG. 2: Imagine fitting two free parameters. Relations are illustrated, between the basis of physical parameters $\{a_0, a_1\}$, the principal axis basis $\{y_0, y_1\}$, and the α_s-excursion basis $\{z_0, z_1\}$.

of the a_1 direction to the CTEQ6.6 uncertainty for an observable X can be evaluated as

$$\Delta X_1^2 = \frac{1}{4}\left(X(B) - X(D)\right)^2, \tag{7}$$

where $X(B)$ and $X(D)$ are the values of X at points B and D in the figure.

Next, consider the minimal and maximal excursions of the parameter a_0 allowed inside the ellipse, for a free parameter a_1. These excursions are reached at points A and C shown in the right inset of Fig. 1. The uncertainty along this direction is

$$\Delta X_0^2 = \frac{1}{4}\left(X(A) - X(C)\right)^2. \tag{8}$$

In practice, points A and C are found by a scan of the dependence of $\Delta\chi^2(a_0, a_1)$ on a_0, for a varying a_1. To describe the uncertainty of an observable X, we rotate the $\{a_0, a_1\}$ basis to a basis $\{y_0, y_1\}$ of eigenvectors of the Hessian matrix. The orthogonal directions $\mathbf{y}_0$ and $\mathbf{y}_1$ specify the principal axes of the ellipse, as seen in the left inset of Fig. 2. Since the maximal range of the a_0 parameter is specified by the line segment AC in Fig. 1 or Fig. 2, we could further rotate the $\{y_0, y_1\}$ basis to another basis $\{z_0, z_1\}$, after rescaling the eigenvectors $\mathbf{y}_0$ and $\mathbf{y}_1$ to unity, as shown in the right inset of Fig. 2. The vector $\mathbf{z}_0$ is chosen to be along the AC direction. It is shown in the Appendix that the vector $\mathbf{z}_1$ is along the direction of the line segment BD. It is perpendicular to the $\mathbf{z}_0$ direction in the $\{z_0, z_1\}$ representation, even though the $\mathbf{z}_0$ and $\mathbf{z}_1$ directions are not perpendicular in the $\{a_0, a_1\}$ representation. Taking this conclusion for granted here, we could easily see from the left inset of Fig. 2 that there is no a_0 dependence along the $\mathbf{z}_1$ direction. From the right inset, the total uncertainty of the observable X can be computed as

$$\Delta X^2 = \frac{1}{4}\left[\left(X(A) - X(C)\right)^2 + \left(X(B) - X(D)\right)^2\right] \tag{9}$$

$$= \Delta X_0^2 + \Delta X_1^2. \tag{10}$$

This uncertainty is equal to the quadrature sum of the uncertainties along the AC and BD directions, as has been stated.

We see that the $\mathbf{z}_0$ and $\mathbf{z}_1$ directions, which are initially not orthogonal in the $\{a_0, a_1\}$ basis, can be made such by rotation and scaling. Alternatively, the orthonormal $\{z_0, z_1\}$ basis can be obtained from the $\{a_0, a_1\}$ basis by a shear transformation in the negative a_1 direction, followed by a scaling transformation along the a_0 direction. In the 23-dimensional case, the condition $a_0 = z_0 = 0$ defines the hyperplane spanned by the 22 CTEQ6.6 PDF parameters, for the fixed best-fit α_s. This hyperplane is made orthogonal to z_0 by a rotation and scaling transformation, or, equivalently, by a shear and scaling transformation, analogously to the 2-dimensional case. See the Appendix for the full discussion.

B. Fits with a variable α_s: explicit realizations

We will now explicitly construct two next-to-leading order (NLO) fits of the kinds described above, using the CTEQ6.6 sample [12] of hadronic data [1]. To this end, we modify

[1] Like in all CTEQ global fits, we include full NLO matrix elements in DIS and most Drell-Yan observables. In calculations for inclusive jet production and W lepton asymmetry, where the full NLO/resummed results

the setup of the CTEQ6.6 analysis to allow $\alpha_s(M_Z)$ to vary within the global fit, and to constrain these variations by the world-average (w.a.) value $(\alpha_s)_{w.a.} \pm (\delta\alpha_s)_{w.a.} = 0.118 \pm 0.002$ (at 90% CL) included as a separate data input, in addition to the complete CTEQ6.6 set of hadronic scattering data. Agreement with this *precision data value* is just as desirable (or more so) as agreement with individual data points in the hadronic data sets. So, to assure this agreement, we add a new contribution $\chi^2_{\alpha_s}$ to the global log-likelihood function in the global analysis:

$$\chi^2 = \chi^2_{\text{CTEQ6.6}} + \chi^2_{\alpha_s}, \tag{11}$$

where

$$\chi^2_{\alpha_s} = \lambda \left[\frac{\alpha_s(M_z) - (\alpha_s(M_z))_{w.a.}}{(\delta\alpha_s)_{w.a.}} \right]^2. \tag{12}$$

The $\chi^2_{\alpha_s}$ term is multiplied by a weighting factor λ to match the confidence interval of the world-average α_s with the tolerance on the increase in χ^2 allowed for acceptable fits. We choose λ so that $\alpha_s(M_z)$ values outside of the 90% CL interval, $\alpha_s(M_Z) \leq 0.116$ or $\alpha_s(M_Z) \geq 0.120$, result in a penalty beyond the tolerance for the increase in χ^2.

Without the contribution of the world-average value, the constraints of the global fit on α_s are relatively weak,

$$\alpha_s(M_Z) = 0.118 \pm 0.005 \qquad (90\%\ \text{CL}). \tag{13}$$

The central value returned by this fit is practically identical to either the PDG or Les Houches workshop central values. When the world-average constraint is included, the final result of the global analysis changes to

$$\alpha_s(M_Z) = 0.1180 \pm 0.0019 \qquad (90\%\ \text{CL}). \tag{14}$$

Again, its central value is practically the same, but the uncertainty is much smaller. Thus the constraint on $\alpha_s(M_Z)$ is dominated by the world-average uncertainty, $(\delta\alpha_s)_{w.a.}$.

We explore the vicinity of the best fit in two ways. First, we construct best-fit PDFs sets for four alternative values of $\alpha_s(M_Z)$,

$$\alpha_s(M_Z) = 0.116, \quad 0.117, \quad 0.119, \text{ and } \quad 0.120. \tag{15}$$

These PDFs are named as

$$\text{CTEQ6.6AS} = AS_{-2}, \quad AS_{-1}, \quad AS_{+1}, \text{ and } \quad AS_{+2}. \tag{16}$$

The CTEQ6.6AS PDFs for the two extreme variations, AS_{-2} and AS_{+2}, correspond to slightly more than two standard deviations according to the PDG error in Eq. (2), or

are prohibitively CPU-extensive, K-factor tables are used to look up the ratio of the NLO cross section to the LO cross section (with the LO cross section calculated using the NLO PDFs, and using the 2-loop α_s) separately for each data point. The look-up tables depend very weakly on the input PDF parameters. They are updated in the course of the fitting to ensure that the K-factors have not drifted from their initial values, and we again check that the calculations retain their full NLO accuracy at the end of the global fit.

approximately to the 90% CL uncertainty according to the 2009 Les Houches prescription in Eq. (3). The intermediate PDF sets, AS_{-1} and AS_{+1}, provide additional information on the α_s dependence.

Alternatively, the combined PDF and α_s uncertainty is quantified by the diagonalization of the Hessian matrix, in terms of 46 extreme PDF eigenvector sets for 23 independent combinations of theoretical parameters. Each eigenvector set is associated with its own $\alpha_s(M_Z)$ value within the best-fit range of Eq. (14). This series of eigenvector PDFs is called CTEQ6.6FAS. It will be compared with the CTEQ6.6AS series in the next section.

III. NUMERICAL RESULTS

A. Comparison of the PDF and α_s uncertainties

The AS_{-2} and AS_{+2} PDFs of the CTEQ6.6AS series for the gluon, u quark, and s quark are compared with the CTEQ6.6 PDF uncertainty band for the momentum scale $Q = 2$ GeV in Fig. 3, and for $Q = 85$ GeV in Fig. 4. Each figure shows the ratio of $f(x, Q)$ in the AS_n set to the CTEQ6.6M PDF of the same flavor (solid and dashed curves), as well as the asymmetric CTEQ6.6 PDF uncertainty for this flavor (shaded region), as functions of the momentum fraction x. The upper and lower boundaries $1 \pm [\Delta f(x, Q)]_{\pm}/f(x, Q)$ of the PDF uncertainty band are given by the asymmetric PDF errors, as

$$
\begin{aligned}
[\Delta f(x,Q)]_+ &= \sqrt{\sum_{i=1}^{22} [f_i(x,Q) - f_0(x,Q)]^2} \quad \text{for} \quad f_i > f_0, \\
[\Delta f(x,Q)]_- &= \sqrt{\sum_{i=1}^{22} [f_i(x,Q) - f_0(x,Q)]^2} \quad \text{for} \quad f_i < f_0.
\end{aligned} \tag{17}
$$

For the shown partons (g, u, and s), the PDF uncertainty is larger than the α_s dependence over the range $0.116 \leq \alpha_s(M_Z) \leq 0.120$. The figures for the d quark (not shown) looks similar to the figures for the u quark. The relation between the PDF and α_s uncertainties for $\bar{u}$ and $\bar{d}$ quarks is in between these relations for the u and s quarks. Finally, the c and b uncertainties are qualitatively similar to those for g.

We now add the CTEQ6.6 PDF uncertainty in quadrature to the CTEQ6.6AS α_s uncertainty, and compare this sum to the uncertainty of the 23-parameter fit CTEQ6.6FAS with the floating α_s and the world average constraint. The result is shown in Fig. 5. The figure compares the PDF uncertainty bands from the two realizations of the α_s fit, for the parton flavors that show the largest α_s dependence: gluons and charm quarks. The CTEQ6.6 error band is indicated by dashed lines; while the two realizations of the α_s fit by a hatched error band with solid borders and a filled error band with dotted borders. We observe only small differences between the CTEQ6.6+CTEQ6.6AS and CTEQ6.6FAS PDFs for the flavors shown in the figure. The differences are even smaller for the other quark PDFs. Thus, the two realizations of the α_s series produce nearly identical results, confirming the adequacy of adding the PDF and α_s uncertainties in quadrature.

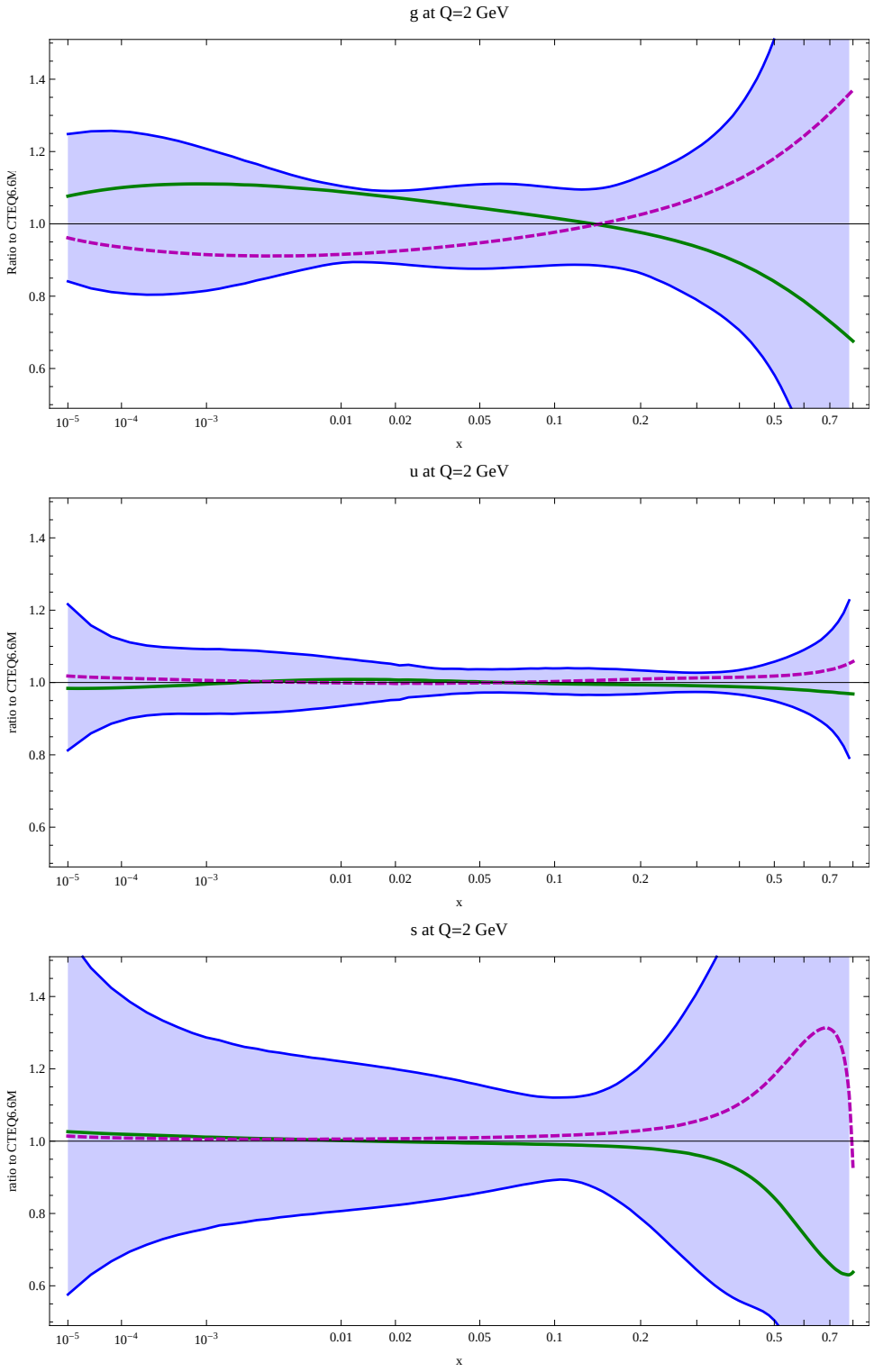

FIG. 3: Comparison of the PDF and α_s uncertainties for the gluon, u quark, and s quark, at $Q = 2\,\mathrm{GeV}$. The vertical axis is the *ratio* of $f(x, Q)$ to the CTEQ6.6M best fit. The CTEQ6.6 PDF uncertainty range for the central α_s value, 0.118, is shown as the shaded region. The ratio of AS_{+2} and AS_{-2} PDFs to the corresponding CTEQ6.6M PDFs are shown as the solid and dashed curves, respectively.

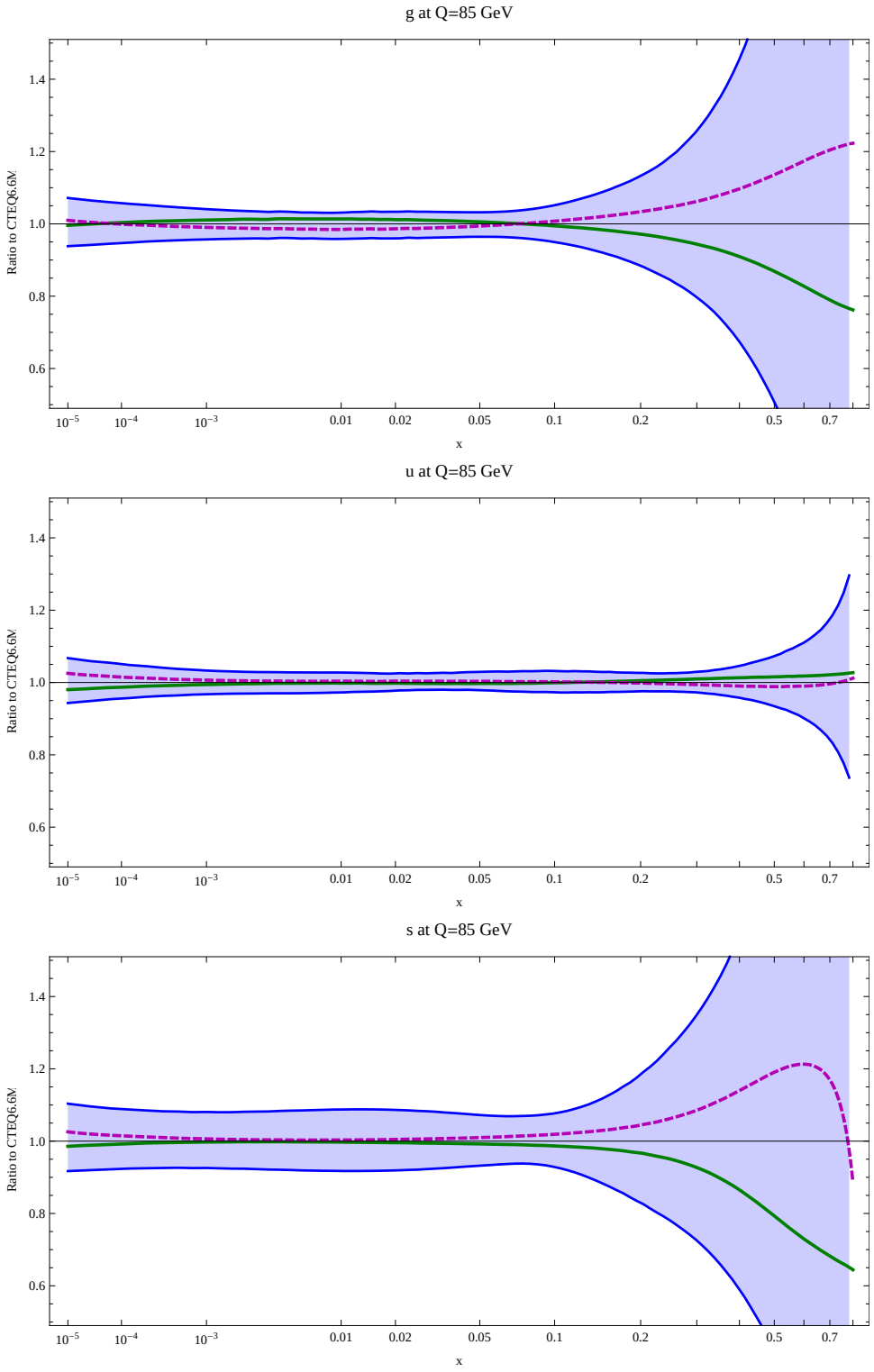

FIG. 4: The same as Fig. 3, for $Q = 85\,\mathrm{GeV}$.

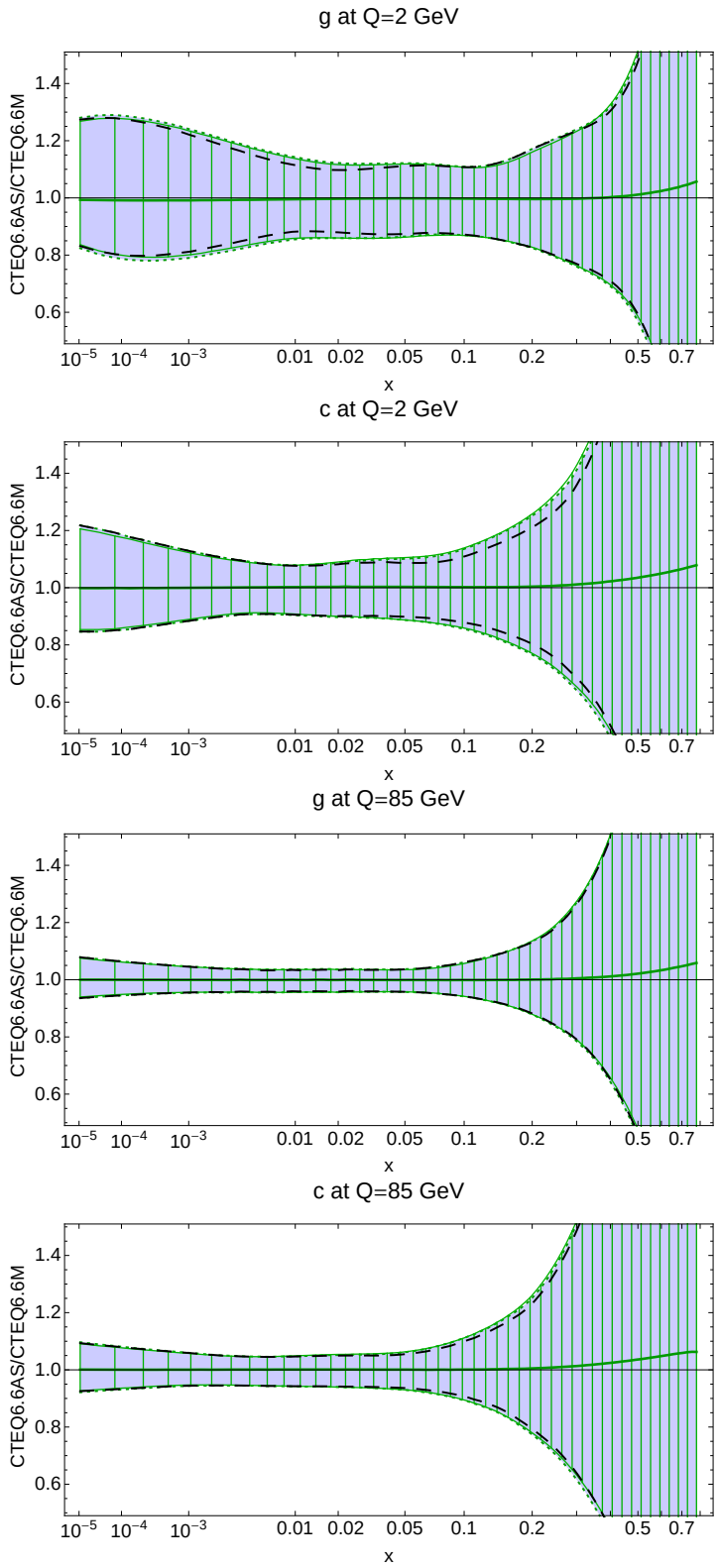

FIG. 5: Comparison of CTEQ6.6+CTEQ6.6AS (blue solid) and CTEQ6.6FAS (green hatched) uncertainty bands and CTEQ6.6 uncertainty bands (indicated by black dashed lines), normalized to the standard CTEQ6.6M fit, for gluon and charm PDFs at $Q = 2$ and 85 GeV.

B. Uncertainties of cross section predictions

Using the α_s series of PDFs, and including the PDF uncertainty, we can estimate the uncertainties of cross section calculations.

For any calculated quantity σ, we denote the central prediction, corresponding to $\alpha_s(M_Z) = 0.118$, by σ_0. There are two contributions to the uncertainty. The symmetric PDF uncertainty, denoted $\Delta\sigma_{\rm PDF}$, is calculated from the Hessian error PDFs by the "master formula" [1]

$$\Delta\sigma_{\rm PDF} = \frac{1}{2}\sqrt{\sum_{i=1}^{d}\left(\sigma_i^{(+)} - \sigma_i^{(-)}\right)^2}\,. \tag{18}$$

Here d is the dimension of the parameter space, and $\sigma_i^{(\pm)}$ is calculated with the eigenvector PDFs.[2] The α_s uncertainty of σ is

$$\Delta\sigma_{\alpha_s} = \frac{1}{2}\sqrt{[\sigma_0(A_{-2}) - \sigma_0(A_2)]^2}. \tag{19}$$

The combined uncertainty $\Delta\sigma$ for CTEQ6.6+CTEQ6.6AS is

$$(\Delta\sigma)^2 = (\Delta\sigma_{\rm PDF})^2 + (\Delta\sigma_{\alpha_s})^2\,. \tag{20}$$

For the CTEQ6.6FAS series, the full uncertainty is computed according to Eq. (18) for $d = 23$. Then the prediction for the quantity σ is

$$\sigma = \sigma_0 \pm \Delta\sigma. \tag{21}$$

Table I lists the predictions for a sample set of NLO cross sections at the Tevatron and LHC energies. Four processes are calculated: $t\bar{t}$ production [18, 19], using a program from Ref. [19]; and Standard Model Higgs boson production, $gg \to H$, for Higgs masses $M_H = 120, 160$ and $250\,{\rm GeV}$ [20]. The production cross sections are predicted for three LHC energies: 7, 10 and 14 TeV. The central predictions σ_0 and uncertainty ranges $\Delta\sigma$ are given first for CTEQ6.6+AS (fitting method 1) and then for CTEQ6.6FAS (fitting method 2). Clearly, the predictions by the two methods are very close, even though not identical due to secondary effects like deviations from the quadratic approximation.

In fitting method 1, we can also compare the relative sizes of the PDF uncertainty and the α_s uncertainty. For processes dominated by the gluon scattering, such as the $t\bar{t}$ production or Higgs production examined here, the α_s and PDF uncertainties can be comparable, as is observed in the Table. These results are illustrated for $t\bar{t}$ production at the LHC with energies 7 TeV and 14 TeV in Fig. 6, and for Higgs boson production ($M_H = 120\,{\rm GeV}$) at the LHC with energies 7 TeV and 14 TeV in Fig. 7. The figures show the cross sections from individual CTEQ6.6(AS) eigenvector sets, as well as the resulting PDF and α_s uncertainties, versus

[2] For these PDFs, based on CTEQ6.6, the number of fitting parameters is $d = 22$. PDF sets $+i$ and $-i$ are variations of the central fit corresponding to displacements in the $+$ and $-$ directions along eigenvector i. The symmetric error formula, Eq. (18), is sufficient for the comparison here; at other times, we would use more detailed asymmetric error estimates, Eq. (17).

Process	CTEQ6.6+CTEQ6.6AS				CTEQ6.6FAS
$t\bar{t}$ (171 GeV)	σ_0	$\Delta\sigma_{PDF}$	$\Delta\sigma_{\alpha_s}$	$\Delta\sigma$	$\sigma_0 \pm \Delta\sigma$
LHC, 7 TeV	157.41	10.97	7.54	13.31	160.10 ± 13.93
LHC, 10 TeV	396.50	18.75	16.10	24.71	400.48 ± 25.74
LHC, 14 TeV	877.19	28.79	30.78	42.15	881.62 ± 44.27
$gg \to H$ (120 GeV)	σ_0	$\Delta\sigma_{PDF}$	$\Delta\sigma_{\alpha_s}$	$\Delta\sigma$	$\sigma_0 \pm \Delta\sigma$
Tevatron, 1.96 TeV	0.63	0.042	0.032	0.053	0.64 ± 0.055
LHC, 7 TeV	10.70	0.31	0.32	0.45	10.70 ± 0.48
LHC, 10 TeV	20.33	0.66	0.56	0.87	20.28 ± 0.93
LHC, 14 TeV	35.75	1.31	0.94	1.61	35.63 ± 1.70
$gg \to H$ (160 GeV)	σ_0	$\Delta\sigma_{PDF}$	$\Delta\sigma_{\alpha_s}$	$\Delta\sigma$	$\sigma_0 \pm \Delta\sigma$
Tevatron, 1.96 TeV	0.26	0.026	0.015	0.030	0.26 ± 0.031
LHC, 7 TeV	5.86	0.16	0.18	0.24	5.88 ± 0.26
LHC, 10 TeV	11.73	0.33	0.33	0.47	11.72 ± 0.50
LHC, 14 TeV	21.48	0.68	0.56	0.88	21.43 ± 0.94
$gg \to H$ (250 GeV)	σ_0	$\Delta\sigma_{PDF}$	$\Delta\sigma_{\alpha_s}$	$\Delta\sigma$	$\sigma_0 \pm \Delta\sigma$
Tevatron, 1.96 TeV	0.055	0.0099	0.0044	0.011	0.058 ± 0.012
LHC, 7 TeV	2.30	0.085	0.081	0.12	2.32 ± 0.12
LHC, 10 TeV	5.08	0.14	0.15	0.21	5.10 ± 0.22
LHC, 14 TeV	10.03	0.26	0.27	0.37	10.04 ± 0.41

TABLE I: Predicted cross sections for $t\bar{t}$ production and Higgs boson production at the Tevatron and LHC, for different scattering energies. Cross sections are given in picobarns. The renormalization and factorization scales are chosen to be the top quark mass and Higgs boson mass for the $t\bar{t}$ and $gg \to H$ cross sections, respectively. The columns present the central prediction σ_0, PDF uncertainty Δ_{PDF}, α_s uncertainty Δ_{α_s}, and total PDF+α_s uncertainty $\Delta\sigma$, computed according to the CTEQ6.6AS and CTEQ6.6FAS methods.

the corresponding $\alpha_s(M_Z)$ values. The overall prediction based on the full CTEQ6.6+AS eigenvector set, $\sigma \pm \Delta\sigma$, is shown by an error bar: the inner bar is the PDF error alone; the outer bar is the combined PDF + α_s error. According to the figures, the α_s uncertainty of the total cross section in the shown processes constitutes between 70% and 110% of the PDF uncertainty.

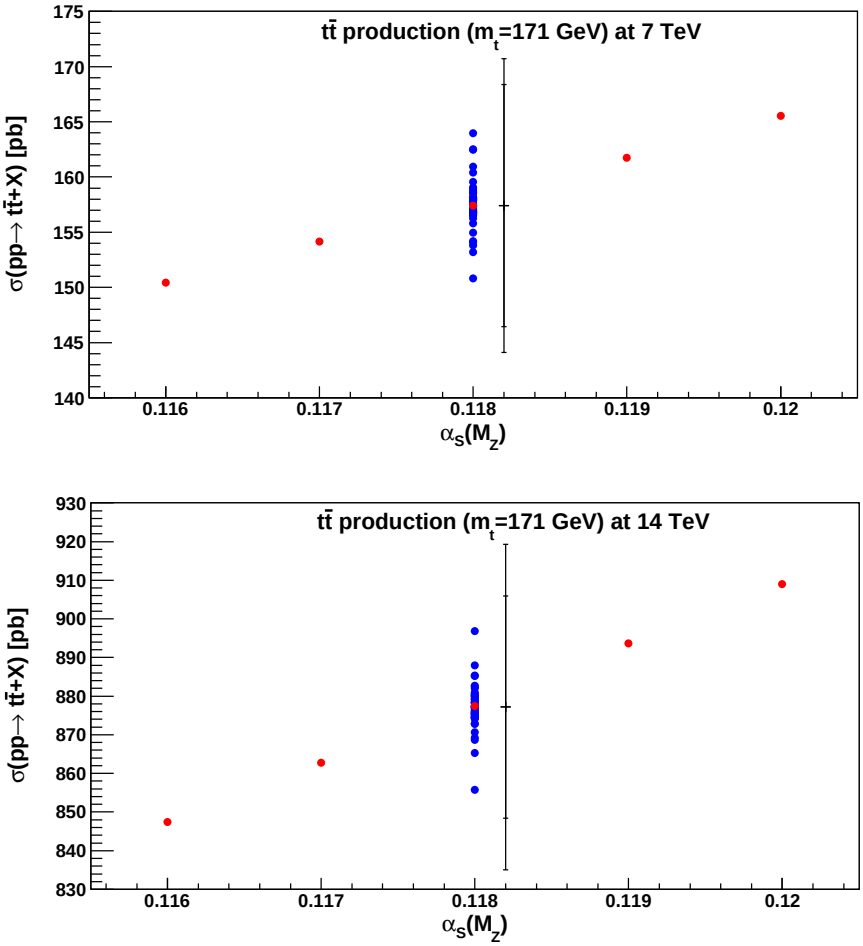

FIG. 6: Cross section for $t\bar{t}$ production at the LHC, at center-of-mass energies of 7 TeV and 14 TeV, as a function of $\alpha_s(M_Z)$. The predictions for the the central set and the 44 eigenvector PDF sets are shown for $\alpha_s(M_Z) = 0.118$. For the other values of $\alpha_s(M_Z)$ (0.116, 0.117, 0.119, 0.120), only the central prediction is shown. The combined uncertainty range (CTEQ6.6+CTEQ6.6AS) is shown as the error bar; cf. Table I. The inner error bar is the PDF error alone, and the outer error bar is the combined PDF + α_s uncertainty.

IV. CORRELATION BETWEEN α_s AND THE PDFS

The independence of the α_s uncertainty from the PDF uncertainty in the CTEQ6.6AS method does not preclude existence of some correlation between the α_s and PDF parameters. This correlation arises from the hadronic scattering experiments, which probe a variety of combinations of the PDFs and α_s. Let us now examine which PDF flavors are most affected by variations in $\alpha_s(M_Z)$, and which scattering experiments impose the most relevant constraints.

For this purpose, we go back to the full fit CTEQ6.6FAS with the floating α_s and examine the correlation between $\alpha_s(M_Z)$ and individual PDF $f_a(x, Q)$ using the method outlined in

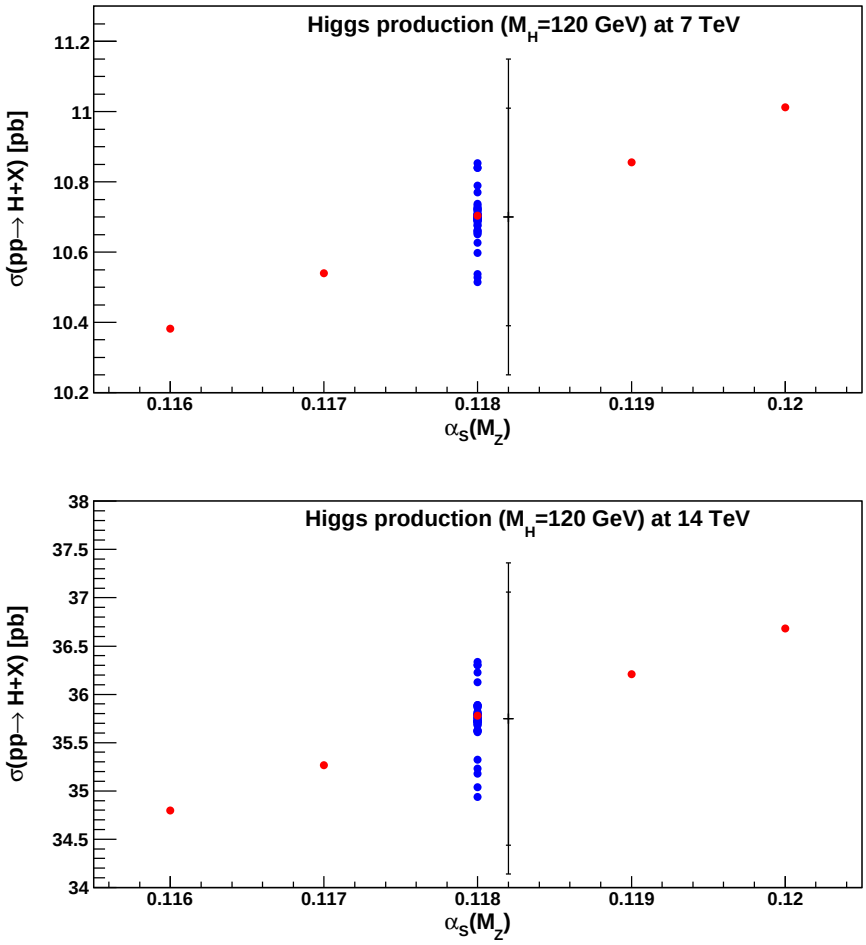

FIG. 7: Same as Fig. 6, for production of Standard Model Higgs boson with mass 120 GeV at the LHC.

Refs. [1, 12, 13]. Let $Y_i^{(\pm)}$ (for $i = 1, ..., 23$) denote one of 46 eigenvector PDFs $f_a(x, Q)$ for a chosen a, x, and Q, corresponding to the maximal acceptable displacements of orthonormal PDF parameters $\{z_1, ..., z_{23}\}$ from their best-fit values in the positive (+) and negative (−) directions, respectively. Given 46 values of $\alpha_s(M_Z)$ (denoted as $X_i^{(\pm)}$), we compute the correlation cosine,

$$\cos\varphi = \frac{1}{4\Delta X\,\Delta Y}\sum_{i=1}^{23}\left(X_i^{(+)} - X_i^{(-)}\right)\left(Y_i^{(+)} - Y_i^{(-)}\right), \tag{22}$$

where ΔX and ΔY are the symmetric PDF errors, found from Eq. (18) for $d = 23$. Values of $\cos\varphi$ close to $+1$, 0, -1 indicate strong correlation, no correlation, and strong anti-correlation between $\alpha_s(M_Z)$ and the PDF $f_a(x, Q)$ in question.

Fig. 8 shows $\cos\varphi$ versus x, for the PDFs that have the largest correlations with $\alpha_s(M_Z)$, at $Q = 2$ and 85 GeV. Other PDFs have little correlation with $\alpha_s(M_Z)$ and are not shown in the figure. We also identify the experimental data sets that cause the largest correlations.

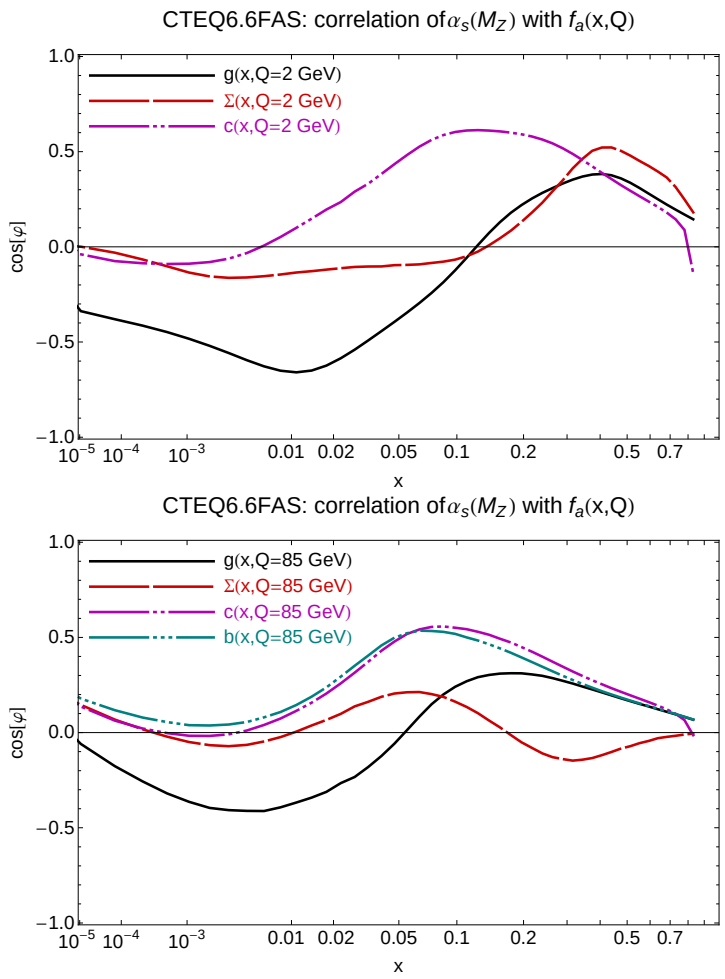

FIG. 8: Correlation cosine $\cos\varphi$ between the best-fit QCD coupling $\alpha_s(M_Z)$ and PDFs $f_a(x,Q)$, plotted as a function of the momentum fraction x at $Q = 2$ GeV and $Q = 85$ GeV.

As shown in Fig. 8, at $Q = 2$ GeV , the most significant correlation (or anticorrelation) of $\alpha_s(M_Z)$ occurs with:

- the gluon PDF $g(x,Q)$ (anticorrelated at $x \sim 0.01$ due to neutral-current DIS constraints from HERA);
- the singlet PDF $\Sigma(x,Q)$ (correlated at $x \sim 0.4$ due to constraints imposed by BCDMS and NMC neutral-current deep-inelastic scattering (DIS) data);
- the heavy-quark PDFs, $c(x,Q)$ and $b(x,Q)$ (correlated at $x = 0.05 - 0.2$ due to constraints by HERA charm and bottom semi-inclusive DIS data).

At $Q = 85$ GeV, these correlations are reduced by the PDF evolution, with the exception of the heavy-quark PDFs. The same conclusions can be drawn from the simpler CTEQ6.6AS analysis, as well as from the correlation analysis by NNPDF [11].

The best-fit value of $\alpha_s(M_Z)$ is thus determined by several types of the data, probing the gluon evolution in DIS at moderately small x, the singlet PDF evolution in DIS at large x,

and HERA charm semi-inclusive DIS data. The correlation of each kind disappears if the relevant data set is removed. For example, it is believed that the low-Q/large-x BCDMS and NMC DIS data prefer somewhat lower $\alpha_s(M_Z)$ values than the rest of the experiments [21], indicating the possible presence of higher-twist terms that are not explicitly included in most PDF analyses [22]. With the most suspect part of these data excluded, the central $\alpha_s(M_Z)$ value indeed increases to 0.119 – 0.120; the spike in $\cos\varphi$ for the singlet PDF $\Sigma(x,Q)$ at $x \sim 0.4$ also disappears from Fig. 8.

V. CONCLUSION

We conclude with this important point: while there are correlations between $\alpha_s(M_Z)$ and some PDFs, as demonstrated in Figure 8, the total uncertainty of the CTEQ6.6FAS fit is essentially the same as the CTEQ6.6+CTEQ6.6AS fit, as demonstrated in Figure 5. The CTEQ6.6+CTEQ6.6AS method captures all the correlation of α_s and PDF parameters as a mathematical consequence of the quadratic approximation, upon which the Hessian method is based. The ability of the quadrature method to reproduce the total PDF+α_s uncertainty (also observed in a related NNPDF study [11]) is thus more than a coincidence.

We may predict cross sections using the CTEQ6.6FAS PDFs, with their 23×2 Hessian eigenvector PDF sets. The results are shown in the final column of Table I. The predictions agree well with those of the simpler method based on the CTEQ6.6+CTEQ6.6AS PDF sets. This again justifies our recommendation to use the CTEQ6.6+CTEQ6.6AS PDFs to estimate uncertainties induced by the PDFs and α_s, according to Eq. (20).

The experimental measurements included in the world average give a smaller α_s uncertainty than the hadronic scattering data can match. Therefore, when making predictions for the LHC or Tevatron colliders, the most realistic estimate of the α_s uncertainty will result from the CTEQ6.6+CTEQ6.6AS α_s series, which covers

$$\alpha_s(M_Z) = 0.116,\ 0.117,\ 0.118,\ 0.119,\ \text{and } 0.120. \tag{23}$$

The central value of this series (assumed in the CTEQ6.6 eigenvector set) is closest to the world average. The PDFs for α_s values at 0.116 and 0.120 provide an estimate of the α_s uncertainty at approximately 90% CL, which is to be combined in quadrature with the CTEQ6.6 PDF uncertainty. The CTEQ6.6AS PDF sets are available from the CTEQ6.6 PDF website [23] and as a part of the LHAPDF library [24].

Acknowledgments
We thank S. Forte, J. Rojo, R. Thorne, A. Vicini, participants of the PDF4LHC and Les Houches 2009 workshops, and members of CTEQ for stimulating discussions. This work was supported in part by the U.S.DOE Early Career Research Award DE-SC0003870; by the U.S. National Science Foundation under grant PHY-0855561; by the National Science Council of Taiwan under grants NSC-98-2112-M-133-002-MY3 and NSC-99-2918-I-133-001; by LHC Theory Initiative Travel Fellowship awarded by the U.S. National Science Foundation under grant PHY-0705862; and by Lightner-Sams Foundation. C.-P. Y. would also like to thank the hospitality of National Center for Theoretical Sciences in Taiwan and Center for High Energy Physics, Peking University, in China, where part of this work was done.

Appendix A: Error matrix diagonalization for the QCD coupling

In this Appendix we provide a proof of the method advocated in this study, whereby an uncertainty induced by the small variation of α_s value is added in quadrature with the fixed-α_s PDF uncertainty. The only necessary assumptions are the usual quadratic approximation for χ^2, and the understanding that the PDF sets for the "up" and "down" variations of α_s are obtained through fits in which all of the PDF parameters are free.

Assume we have an initial fit in which there are N free parameters, with an additional parameter held fixed. The additional parameter can be any additional degree of freedom that one wishes to study. Here it is intended to be $\alpha_s(M_Z)$; but no special features of that parameter will be invoked.

We begin by carrying out the traditional Hessian diagonalization procedure, applicable in the vicinity of the best-fit (BF) combination of the PDF parameters. Starting from the "shape" parameters that control the input PDFs at scale Q_0, we define new fitting parameters $a_1, \ldots, a_N$ that are coefficients of the eigenvectors of the original Hessian matrix, to obtain

$$\chi^2 \,=\, \chi^2_{\rm BF} \,+\, \sum_{i=1}^{N} a_i^2 \;. \tag{A1}$$

We now wish to include an additional degree of freedom in the fit. Hence we introduce a new parameter a_0, which could be defined as proportional to the deviation of $\alpha_s(M_Z)$ from its best-fit value, $\alpha_s(M_Z) - \alpha_{s,\rm BF}(M_Z)$; or better, as proportional to $\ln[\alpha_s(M_Z)/\alpha_{s,\rm BF}(M_Z)]$. Including the new degree of freedom, we have, in general,

$$\chi^2 \,=\, \chi^2_{\rm BF} \,+\, b\, a_0 \,+\, \sum_{i=0}^{N}\sum_{j=0}^{N} H_{ij}\, a_i\, a_j \;, \tag{A2}$$

where

$$\mathbf{H} \,=\, \begin{pmatrix} 1 & p_1 & p_2 & p_3 & \ldots & p_N \\ p_1 & 1 & 0 & 0 & \ldots & 0 \\ p_2 & 0 & 1 & 0 & \ldots & 0 \\ p_3 & 0 & 0 & 1 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \ldots & \cdot \\ p_N & 0 & 0 & 0 & \ldots & 1 \end{pmatrix} \;. \tag{A3}$$

Here

$$p_i = \frac{1}{2}\left(\frac{\partial^2\chi^2}{\partial a_0 \partial a_i}\right)_{\rm BF} \;. \tag{A4}$$

$\mathbf{H}$ describes the correlation between a_0 and a_i, as it gives the variation of χ^2 in the $\{a_0, a_i\}$-subspace. The linear term $b\, a_0$ in (A2) allows for the possibility that the new minimum might not be at $a_0 = 0$. Note that $b = (\partial\chi^2/\partial a_0)_{\rm BF}$, which describes how χ^2 varies with a_0 at the original best fit. The new diagonal element H_{00} was chosen to be 1 by including an appropriate scaling factor in the definition of a_0.

As a_0 varies, the PDF parameters $a_1, ..., a_N$ are adjusted so that to reduce the increase in χ^2. The direction along which this increase is minimal (direction AC in Fig. 1) corresponds to $\partial\chi^2/\partial a_i = 0$ for $i = 1, ...N$, *i.e.*, it is a line given by $a_i = -p_i a_0$ according to Eqs. (A2) and (A3).

Define new variables z_i as

$$z_i \;=\; a_i + p_i\, a_0\,, \quad \text{for } i = 1, ..., N, \tag{A5}$$

$$z_0 \;=\; \sqrt{1-C^2}\, a_0 + \frac{b}{2\sqrt{1-C^2}}, \tag{A6}$$

where

$$C = \sqrt{\sum_{i=1}^{N} p_i^2}. \tag{A7}$$

Since χ^2 is assumed to have a minimum, hence, $0 \leq C < 1$. The transformation (A5, A6) to the z_i coordinates consists of an a_0-dependent translation (shear) along the PDF coordinates a_i, followed by a scaling and a translation along the α_s coordinate a_0. In the new coordinates, χ^2 is diagonal,

$$\chi^2 \;=\; \chi^2_{\rm BF} \;-\; \frac{b^2}{4(1-C^2)} \;+\; \sum_{i=0}^{N} z_i^2. \tag{A8}$$

Variations along z_0 and z_i for $i = 1, ...N$ are explicitly independent, and we can calculate the symmetric uncertainty ΔX for an observable $X(z_0, z_1, ..., z_N)$ by

$$(\Delta X)^2 \;=\; \frac{1}{4}\sum_{i=0}^{N} \left[X(0, ..., z_i = T, ..., 0) - X(0, ..., z_i = -T, ..., 0)\right]^2, \tag{A9}$$

for tolerance $\Delta\chi^2 = T^2$.

Below, we will focus on the case that $b = 0$, which is valid in this paper. From Eqs. (A5) and (A6), the extreme values of a_i consistent with $\Delta\chi^2 = T^2$ are $a_i^{\pm} = z_i^{\pm} = \pm T$. They occur at $a_0 = z_0 = 0$ (*i.e.*, at the best-fit α_s) and $z_k = 0$ for $k \neq i$. The extreme values of a_0 occur at $z_0^{\pm} = \pm T$ and $z_i = 0$ for $i = 1, ..., N$, *i.e.*, $a_0^{\pm} = \pm T/\sqrt{1-C^2}$ and $a_i^{\pm} = -p_i a_0^{\pm}$. Hence Eq. (A9) corresponds exactly to adding in quadrature the uncertainty based on the allowed range of a_i computed at $a_0 = 0$ (analogous to the CTEQ6.6 uncertainty in our application) to the uncertainty based on the allowed range of a_0, computed with a_i at their preferred values for the extreme excursions in a_0 (analogous to the CTEQ6.6AS uncertainty).

An alternative way to construct the z_i basis is to find the eigenvectors of the $(N+1)\times(N+1)$ Hessian matrix $\mathbf{H}$ in Eq. (A3) and then perform an additional orthogonal transformation. To find the eigenvectors, note that $\mathbf{H}$ can be written in a block form

$$\mathbf{H} = \begin{pmatrix} 1 & \mathbf{p}^T \\ \mathbf{p} & \mathbf{I} \end{pmatrix} \tag{A10}$$

where 1 is a unit element, $\mathbf{p}^T$ is a row vector $(p_1, p_2, ..., p_N)$, and $\mathbf{I}$ is an $N \times N$ unit matrix.

- One eigenvector is

$$\mathbf{V}^{(0)} = F \begin{pmatrix} C \\ \mathbf{p} \end{pmatrix}, \tag{A11}$$

  corresponding to an eigenvalue $\lambda_0 = 1 + C$. The normalization factor is $F = 1/(\sqrt{2}C)$.

- A second eigenvector is

$$\mathbf{V}^{(1)} = F \begin{pmatrix} -C \\ \mathbf{p} \end{pmatrix}, \tag{A12}$$

with an eigenvalue $\lambda_1 = 1 - C$.

- Now consider vectors of the form

$$\mathbf{V} = \begin{pmatrix} 0 \\ \mathbf{q} \end{pmatrix} \tag{A13}$$

where $\mathbf{p}^T\mathbf{q} = 0$. These are orthogonal to the first two eigenvectors; the N-dimensional vector $\mathbf{q}$ is orthogonal to $\mathbf{p}$. For any $\mathbf{q}$, we have $\mathbf{H}\,\mathbf{V} = \mathbf{V}$. All their eigenvalues are $\lambda_j = 1$, for $j = 2, 3, \ldots, N$.

So we have a set of eigenvectors of $\mathbf{H}$, denoted by $\mathbf{V}^{(0)}, \mathbf{V}^{(1)}, \mathbf{V}^{(2)}, \ldots, \mathbf{V}^{(N)}$, and satisfying the orthonormality condition

$$\sum_{k=0}^{N} V_k^{(i)} V_k^{(j)} = \delta_{ij}. \tag{A14}$$

The next step is to use the eigenvectors $\mathbf{V}^{(j)}$ to define new coordinates $t_0, \ldots, t_N$ by

$$a_i \;=\; \sum_{j=0}^{N} t_j \, V_i^{(j)} \,. \tag{A15}$$

It is easy to compute χ^2 from Eq. (A2) by making use of the eigenvector properties

$$\mathbf{H}\,\mathbf{V}^{(j)} \;=\; \lambda_j \,\mathbf{V}^{(j)} \,, \tag{A16}$$

where $\lambda_0 = 1 + C$, $\lambda_1 = 1 - C$, and $\lambda_j = 1$ for $j = 2, \ldots, N$; and then using the orthogonality property (A14). The result has the expected diagonal form

$$\chi^2 \;=\; \chi^2_{\rm BF} \;+\; (1+C) t_0^2 \;+\; (1-C) t_1^2 \;+\; \sum_{i=2}^{N} t_i^2 \,, \tag{A17}$$

which is further simplified to

$$\chi^2 \;=\; \chi^2_{\rm BF} \;+\; \sum_{i=0}^{N} y_i^2 \tag{A18}$$

by rescaling of the first two coordinates:

$$y_0 = \sqrt{1+C}\, t_0; \quad y_1 = \sqrt{1-C}\, t_1; \quad y_i = t_i \text{ for } i = 2, \ldots, N \,. \tag{A19}$$

The function χ^2 is explicitly diagonal in the new coordinates y_i, but not in the parameter a_0 that is of interest here, given that

$$a_0 \;=\; y_0/\sqrt{2(1+C)} \;-\; y_1/\sqrt{2(1-C)} \,. \tag{A20}$$

It can be made so by a further orthogonal transformation called the Data Set Diagonalization [17]. The general method for finding the necessary transformation is to write $a_0^2 = \sum_{i=0}^{N} \sum_{j=0}^{N} G_{ij}\, w_i\, w_j$ and use the eigenvectors of $\mathbf{G}$ as new basis vectors. However, in the present case, the required transformation can be found more simply by observing that one of the new coordinates must be proportional to a_0, and the others must be orthogonal to it. The result is

$$\begin{aligned} z_0 &= \sqrt{\tfrac{1-C}{2}}\, y_0 - \sqrt{\tfrac{1+C}{2}}\, y_1, \\ z_1 &= \sqrt{\tfrac{1+C}{2}}\, y_0 + \sqrt{\tfrac{1-C}{2}}\, y_1, \\ z_j &= y_j \quad \text{for } j = 2, \ldots, N\ , \end{aligned} \tag{A21}$$

which yields

$$\chi^2 = \chi^2_{\rm BF} + \sum_{j=0}^{N} z_j^2, \tag{A22}$$

equivalent to Eq. (A8) with $b = 0$. Because χ^2 is a sum of independent terms for each z_i, the uncertainties of these variables should be combined in quadrature. This conclusion does not depend on the value of C, *i.e.*, the strength of correlation between the new degree of freedom a_0 and the original parameter space of a_i in Eq.(A1). The dependence of χ^2 on a_0 comes entirely from the coordinate z_0. The space spanned by $z_1, \ldots, z_N$ at $z_0 = 0$ has $a_0 = 0$, so it is the same fitting space spanned by $a_1, \ldots, a_N$ in the original N-parameter fit. Hence the uncertainty associated with $z_1, \ldots, z_N$ is the same as the uncertainty associated with $a_1, \ldots, a_N$. It is therefore not necessary to explicitly carry out the transformations, and the overall formula for the uncertainty is the result we have used in Eq. (20).

[1] D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, H.-L. Lai, W.-K. Tung, Phys. Rev. D**65**, 014012 (2001); J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H.-L. Lai, W.-K. Tung, Phys. Rev. D**65**, 014013 (2001).

[2] A. D. Martin, R. G. Roberts, W. J. Stirling, R. S. Thorne, Eur. Phys. J. C**28**, 455 (2003); Eur. Phys. J. C**35**, 325 (2004).

[3] A. D. Martin, W. J. Stirling, R. S. Thorne, G. Watt, Eur. Phys. J. C**63**, 189 (2009).

[4] A. M. Cooper-Sarkar, Conference on Advanced Statistical Techniques in Particle Physics, March 18-22 2002, Durham, UK; J. Phys. G**28**, 2669 (2002).

[5] S. I. Alekhin, Phys. Rev. D**68**, 014002 (2003).

[6] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, A. Piccione, J. Rojo, M. Ubiali, Nucl. Phys. B**809**, 1 (2009).

[7] J. Pumplin, A. Belyaev, J. Huston, D. Stump, W.-K. Tung, JHEP **0602**, 032 (2006).

[8] S. Bethke, Eur. Phys. J. C**64**, 689 (2009).

[9] C. Amsler, et al., Phys. Lett. B**667**, 1 (2008).

[10] A. D. Martin, W. J. Stirling, R. S. Thorne, G. Watt, Eur. Phys. J. C**64**, 653 (2009).

[11] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, M. Ubiali and A. Vicini, in [14]; F. Demartin, S. Forte, E. Mariani, J. Rojo and A. Vicini, arXiv:1004.0962 [hep-ph].

[12] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, D. Stump, W.-K. Tung, C.-P. Yuan, Phys. Rev. D**78**, 013004 (2008).

[13] P. M. Nadolsky and Z. Sullivan, hep-ph/0110378 (2001).
[14] J. R. Andersen et al., *The SM and NLO Multileg Working Group: Summary report*, in Proceedings of the 6th Les Houches Workshop: Physics at TeV Colliders, Les Houches, France, 8-26 Jun 2009, arXiv:1003.1241 [hep-ph].
[15] https://wiki.terascale.de/index.php/PDF4LHC_WIKI.
[16] The H1, ZEUS collaborations, JHEP **1001**, 109 (2010); S. Glazov, arXiv:0911.0159 [hep-ex]; C. Gwenlan, arXiv:0907.3483 [hep-ex].
[17] J. Pumplin, Phys. Rev. D **80**, 034002 (2009); J. Pumplin, Phys. Rev. D **81**, 074010 (2010).
[18] P. Nason, S. Dawson and R. K. Ellis, Nucl. Phys. B **303**, 607 (1988); Nucl. Phys. B **327**, 49 (1989) [Erratum-ibid. B **335**, 260 (1990)].
[19] W. Beenakker, H. Kuijf, W. L. van Neerven and J. Smith, Phys. Rev. D **40**, 54 (1989); W. Beenakker, W. L. van Neerven, R. Meng, G. A. Schuler and J. Smith, Nucl. Phys. B **351**, 507 (1991).
[20] M. Spira, arXiv:hep-ph/9510347.
[21] M. Virchaux and A. Milsztajn, Phys. Lett. B**274**, 221 (1992); S. I. Alekhin, Phys. Rev. D**59**, 114016 (1999); S. I. Alekhin, JHEP **0302**, 015 (2003).
[22] A. Accardi, M. E. Christy, C. E. Keppel, P. Monaghan, W. Melnitchouk, J. G. Morfin, and J. F. Owens, Phys. Rev. D**81**, 034016 (2010); J. Pumplin, Phys. Rev. D**81**, 074010 (2010).
[23] http://hep.pa.msu.edu/cteq/public/cteq6.html.
[24] http://projects.hepforge.org/lhapdf/.